

DOCUMENT RESUME

ED 135 485

PS 009 122

AUTHOR Honig, Alice S.; And Others
TITLE Effects of Testing Style on Language Scores of Four-Year-Old Low-Income "Control" Children in Intervention Projects.
PUB DATE Apr 75
NOTE 30p.; Paper presented at the Biennial Meeting of the Society for Research in Child Development (Denver, Colorado, April 10-13, 1975)
EDRS PRICE MF-\$0.83 HC-\$2.06 Plus Postage.
DESCRIPTORS Culture Free Tests; Educationally Disadvantaged; Enrichment Programs; *Intelligence Tests; Intervention; Low Income Groups; Preschool Children; *Preschool Education; Research; *Socioeconomic Influences; *Testing Problems; *Test Reliability
IDENTIFIERS Illinois Test of Psycholinguistic Abilities; *Testing Styles

ABSTRACT

This study tests the hypothesis that optimum testing style will result in more efficient test performance by 4-year-olds, thus diminishing differences between scores of lower class children who have and have not attended enrichment programs. A review of the literature discusses the effects of testing situation variables, language differences, dialects, ethnicity, motivational factors, and manipulation of pretest conditions on the scores of low income children. In the present study, scores on the Illinois Test of Psycholinguistic Ability (ITPA) were compared, since it was assumed that language scores were particularly likely to reflect the effects of optimum testing style. Data included the ITPA scores of (1) children participating in the Family Development Research Program (FDRP), an enrichment program for infants to school age children from low income families; (2) carefully matched control children who did not participate in the FDRP program but who were tested under FDRP conditions; and (3) children who had served as controls for other enrichment programs, in which testing was not done under FDRP conditions. Children in the FDRP experimental and control groups were not tested until they were at ease in the testing situation, with all attempts made to optimize test conditions. Although FDRP control children scored below children in the FDRP enrichment program, they performed only slightly below national norms on three of the ITPA subtests used in the comparison, and considerably above national norms on the other four subtests. Control children from the other enrichment programs generally performed below national norms on all subtests for which comparable data were available. Results and implications for testing and intervention programs are discussed. (SE)

Effects of Testing Style on Language Scores of Four-Year-Old

Low-Income "Control" Children in Intervention Projects

Alice S Honig, J.R. Lally, L. Webb, F. Armstrong, & N.S. Wright.

The Problem

The purpose of this paper is to raise questions about psychological testing styles. Two conditions can possibly have been created by present testing techniques. The first condition is the possible artificial restriction of scores for low income control populations because of limited testing techniques. The second condition is that estimates of success of enrichment programs may be based on meaningless differences found between a control group with artificially restricted scores and an experimental group with normal scores - thus propagating enrichment techniques that might be of little value.

Test manuals often give sage advice about building rapport with young children and the importance of such rapport in order to insure optimal test scores for any youngster. For example, Cattell (1960) has urged that the following test conditions be adhered to as closely as possible in the testing of young children.

The child should not be sick, tired, sleepy or in an antagonistic or unhappy mood when tested. The confidence of young children cannot be gained by verbal explanations but must be built up through an easy confident manner. Disapproval of a child's actions can invalidate the score on his test. Testers should always be alert to boredom and counteract it with praise, encouragement or quick surprise presentation of new toys. An introductory toy should be used that will insure success but still arouse interest (pp. 74-75).

However, there has not always been sufficient adherence to such precepts and the problem of the effect of testing styles on scores has not received as much attention as it warrants. Testing style is a particularly critical issue to consider when one regards the large number of intervention projects which use scores from contrast or control groups in order to demonstrate gains of experimental

groups. Often the control group scores reported are well below national norms. Enrichment group children performing at the norms are then often considered as "superior" in functioning as a consequence of program input. But are these differences valid? Many researchers (Labov, 19 ; Baratz & Baratz, 1970, and Valentine, 19) have argued the cultural inappropriateness of some tests. But other issues also need to be considered. Lally (1970, 1972) has reported that within the past few years, over one hundred researchers engaged in intervention programs with children from zero to three years formed a group which met three to four times a year with hopes that these meetings would help them better understand each other's work. It was discovered early in the dialogue that the standardized test procedures recommended on most nationally recognized tests were not sufficient to ensure comparable administration from site to site. Suggestions were made to remedy the situation. For example, a videotape of each test battery would be agreed upon by the group and then sent from site to site to ensure quality control. Unfortunately, this remedy and others, although sorely needed, were not acted upon. The meaning of test scores at different sites is still difficult to judge and specific comparisons of test results across sites are hard to make without additional information about testing styles and testing conditions.

Testing Situation Variables and Children's Scores

Inquiry into testing variables which may influence children's scores and thus obscure sought-for relationships between program input variables and child outcome measures have focused on variables aside from those obviously associated with adequate knowledge of test content or of testing procedures with young children. Indeed, inept abilities of testers to gain rapport with young children have long been found to influence the kind and validity of data obtained not only in testing situations but in interviews as well (Yarrow, 1960).

Language differences, dialects, and ethnicity. When the child's commonly used home language or dialect differs from the standard language in which a test is presented, then different degrees of handicap may accrue in test scores for

the children. Nedler & Sebera (1971) have demonstrated this by giving the PPVT IQ test (Peabody Picture Vocabulary Test) both in Spanish and in English to three-year-old bicultural children. As much as a 28 IQ point different in scores was obtained in favor of the children's performance on the Spanish version in contrast to the English form of the PPVT..

Dialect differences, on the other hand, have been found to have more variable effects on test scores. Quay (1970) had black male testers administer the Stanford-Binet IQ test either in standard English or in black dialect to four-and-one-half-year-old Project Head Start children, under differential conditions of reinforcement--either tangible candy reward or praise. It was expected that dialect plus the candy reward should optimize test performance of the children. No differences were found for any of the conditions. That is, intelligence scores, which ranged from 95.4 to 97.1, were affected neither by the dialect used by the examiner nor by the type of reinforcement presented. Quay suggests that different dialects may pose more of a problem for speech production tests but that children who speak a dialect predominantly and also have much exposure to standard English models seemed to have no difficulty in comprehending either the dialect or standard English spoken by the testers.

With respect to speech production research, Resnick, Weld, and Lally (1969) have found that the standardized testing situation markedly reduced the frequency of expressive language among disadvantaged black infants. The total number of toddler vocalizations and words spoken was much higher in the pre-testing situation when mother and infant were alone in the testing room with many toys than in the actual testing situation.

Motivational factors and test scores. Zigler and his colleagues have examined the effects of motivational variables on test scores in a series of papers designed to show the vulnerability of cognitive performance to such factors (Butterfield & Zigler, 1970; Zigler & Butterfield, 1968; Zigler, Hogden, & Stevenson, 1958). Some of this work (Zigler, Butterfield, & Capabianco, 1970) focused on the detrimental

effects of ordinary test situations on retarded children starved for positive social contacts. "It appears that the retarded child's atypically high need for social reinforcement competes in the testing situation with the production of correct responses and, thus, results in intelligence test performance lower than that dictated by the child's intellectual resources. To the extent that the child's need for social reinforcement is ameliorated, this leads one to expect an increase in IQ" (p. 261).

Zigler and his colleagues (1973) later investigated and demarcated motivational factors which led to the less adequate performance of economically disadvantaged children on the PPVT. Their motivational hypothesis contained the assumption that children from poor families are much more wary of a tester than are nondisadvantaged children. Their prediction that a sample of four-and-one-half-year-old children enrolled in Head Start Centers would increase their Peabody IQ scores far more on retest within a week or two than would a contrast group of middle and upper-middle-income children was confirmed. Thus, increased comfort deriving from familiarization with tester seemed critical to the significant positive IQ test-retest score gains of the disadvantaged children in comparison to their contrast group where no such gains accrued from first to second test. Later this study was repeated with additional conditions, one of which was a play session added prior to initial testing in order to allay situational anxiety. The authors found that when

situational wariness is overcome, as by a pleasant play period, the initial IQ score becomes significantly higher than when no effort is made to dissipate the child's general anxiety. The finding that disadvantaged children's performance, as compared with that of nondisadvantaged children, was more influenced by the nature of their interaction with the E is consistent with earlier findings that socially deprived children are more sensitive to variation in the emotional climate created by adults than are nondeprived children" (p. 301).

Connors & Eisenberg (1966) have also reported that Peabody IQ gains achieved by children in a summer Head Start program reflected motivational characteristics of children rather than superior cognitive functioning. Sacks (1952) has reported that Stanford-Binet intelligence scores varied from test to test as a function of the experimentally established social relationship between child and examiner.

Manipulation of pretest conditions. Sensitivity to the importance of children's comfort as a factor influencing test scores has led some researchers in recent years to create conditions which will increase the familiarity and confidence with which children enter the testing situation. Garber (1975) has reported for the Milwaukee Project (which provided intensive enrichment experience for low-income black infants whose mothers' mean IQ scores was below 75) that every effort was made to ensure that infants who were earlier randomly assigned to the control group were comfortable and at ease prior to being tested.

Palmer (1970) was more specific about his methods. He involved black male toddlers in either a carefully sequenced concept training curriculum or in a 'discovery' curriculum to teach concepts such as "hot-cold." He has clearly expressed the philosophy of establishing optimal test conditions, a philosophy which indeed may require much staff time and effort.

No child began assessment until he had met the following criteria for adaptation to the Center's environs: (a) mother out of the testing cubicle, door closed, and child playing with examiner for at least 20 minutes; (b) examiner judged the child was ready for testing or, if uncertain, requested a senior examiner to confirm or disconfirm her opinion.

Most children were sufficiently adapted to the Center by their third visit to meet the above criteria, but some required as many as five visits before testing was begun. If during the assessment the child at any time appeared to be distracted beyond the efforts of the examiner to hold his attention, testing was discontinued until the following visit. Children progressed through the measures at their own pace and required from 4 to 15 hours to complete the battery as

a function of the characteristics of the individual child and the age of testing. (p. 3)

Palmer attributes the scarcity of any socio-economic differences on cognitive test scores and measure to be found among his groups (at 3 years, 8 months of age) to his concern for and attention to optimal testing conditions for all children. He concluded that "It is possible that middle-class children are more adapted to conditions where such testing ordinarily occurs and consequently perform better in those situations, but that the lower-class child when given the opportunity to adapt performs as well" (p. 9).

Sometimes the manipulation of pretest conditions to maximize children's comfort unexpectedly points up the vulnerability of test scores to such manipulations. Costello & Martin (1969) carried out a language enrichment pre-school program for black four-and-one-half year olds in a low-income housing project in Chicago over a period of several years. The first two "waves" of children received the Stanford-Binet IQ test shortly after admission to the program. Wave III children in the third project year were not pretested until 6 weeks after the program was under way. This initial adjustment period had a very favorable effect on scores achieved. For example, the mean IQ pretest score for Wave I children was 88.1; for Wave III children the mean pretest score was 94.7. Wave III children had been carefully chosen to represent an even more disadvantaged group of children than those in Waves I and II. Yet, this difference in scores favored Wave III children who were given an initial adjustment period prior to administration of the pretest. Costello & Martin (1969) observed that:

Although we had selected a larger number of children of low competence level than occurred in the unbiased samplings of Waves I and II, our

observation as well as research evidence suggested that the positive influences of testing procedures on the means would be greater than the negative influence of the increased proportion of low competence subjects. . . . The more supportive test procedures of Wave III shifted the score distribution most favorably for middle scoring children, less so for high scorers and least for low scorers. (p. 13)

Such findings have been very important in increasing our awareness and concern with test conditions, particularly for initial conditions under which pretest experimental group data is gathered for a program and for the testing situation and climate which exist for the "control group" child whenever he or she is tested.

Hypotheses

We have hypothesized in the present study that differences in testing styles which optimize conditions will result in more children performing more effectively and thus tend to decrease the probability of very low scores among low-income control group children. Thus the probability of finding large differences between enrichment groups and their controls should be minimized compared with the low scores typically reported for children from poor socio-economic backgrounds who have not participated in an enrichment program.

We have further hypothesized that since language production has been found particularly vulnerable to comfort and familiarity factors, then an optimal testing style applied to a language test might more effectively reveal the positive influence of this optimal style on control children's scores.

Method

The testing philosophy in the Family Development Research Program (FDRP) (Lally, 1974)--an enrichment program from early infancy to school age for low-income families--has been firmly established as one which optimizes the testing conditions as much as is humanly and creatively possible both for children in the project and for their carefully matched controls from low-income families. No child is tested unless and until he or she appears very comfortable and is at ease in the testing situation. This may necessitate feeding children and scheduling several visits in order to begin and to complete an assessment battery. Seven subtests of the Illinois Test of Psycholinguistic Ability (ITPA) (Kirk, McCarthy, & Kirk, 1968) are administered both to the children in the experimental program and to their longitudinal controls from age four onward. Thus, neither group has been exposed to the ITPA prior to this assessment. It should be made clear that scoring criteria for the ITPA have not been altered in any way, but on the contrary have been adhered to scrupulously. Table 1 provides a description of the ITPA subtests.

 Insert Table 1 about here

Subjects

In order to contrast scores of control group four-year-olds from the Family Development Research Program with scores of comparable low-income peers, we have used ITPA data from a variety of program sources. Some of these programs used the 1961 version of the ITPA. Since the content validity, characteristic items, and standardization characteristics of subtests with earlier and revised versions are exceedingly close, the 1968 psycholinguistic

age norms and scaled score norms derived from raw scores have been used. When project reports only provided one set of scores, the ITPA norm tables permitted a comparison to be made among the three types of scores for all projects regardless of the original form in which the data were presented.

The data come from the Family Development Research Program (FDRP) as well as other projects with varying numbers of children tested at comparable ages: (1) 16 children from the Durham Education Improvement Program (EIP) (Anastasiow, Stedman, & Spaulding, 1968), an educational intervention program designed for black and for white children living in poverty; (2) 101 children from the Karnes et al. (1969) language intervention project, in which low-income children, both black and white, were given subtests of the Illinois Test of Psycholinguistic Ability at age four just prior to participation in five different intervention programs; (3) 18 black children at four-and-one-half years of age from the control group of the Milwaukee Project (Heber, Garber, Harrington, Hoffman, & Falender, 1972); (4) 20 disadvantaged four-and-one-half black children in Sprigle's Learning-to-Learn Project (Van de Riet, 1972); (5) 15 disadvantaged four-and-one-half-year-old children in the Bereiter-Engelmann patterned drill language intervention program (1966) who received two ITPA subtests prior to admission into the program; (6) 132 children from the Horner Preschool Program for black four-and-one-half-year-old children from a low-income housing project in Chicago (Postello & Martin, 1969).

Data from the children in the contrast programs have been compared with the ITPA subtest scores from the Family Development Research Program where optimal testing is the rule. Brief background information for some of the control children in the FDRP project provided in the Appendix should dispel any tendency of readers to assume that the FDRP control children come from more advantaged backgrounds than the other groups.

Results

The results of these comparisons are shown in Table 2. Examination of the scaled ITPA scores available for children in the contrast programs

 Insert Table 2 about here

indicates that children in the control group of the FDRP project performed only slightly below national norms on three of the ITPA subtests used in the comparison, and considerably above the national norms on the other four subtests. Children in the contrast groups performed for the most part below the national norms on all subtests for which comparable data were available.

Children who have not undergone enrichment procedures in the comparison groups are seen to have performed more poorly than FDRP control group four-year-olds. The Durham Program controls performed below national norms on six subtests and slightly above these norms on three subtests. Children from the Karnes programs at age four, prior to enrichment program entry, performed below national norms on five subtests and at the norm on one subtest. In the Milwaukee Project, controls performed below national norms on eight subtests, above on two. In the Sprigle program, controls performed uniformly below national norms on the four subtests given. In the Bereiter-Engelmann program, controls performed well below national norms on the three subtests given. In the Horner Preschool program, controls performed markedly below national norms on both subtests administered.

Figure 1 graphically demonstrates the superior functioning of FDRP control children on ITPA subtests. The scaled scores of these children are seen to lie near or above the mean scaled scores of 35, which has been

net as the standardized national norm for each age group. The bar graph

Insert Figure 1 about here

Indicates a marked tendency for control test scores from other projects to fall at or well below this national scaled score mean of 35.

Figure 2 contrasts scaled scores of FDRP control group children with

Insert Figure 2 about here

scores achieved by experimental children in all the comparison programs including the FDRP program. Within the FDRP program, in each case, the scores of controls fell below the scores of experimental four-year-olds. Yet if we examine FDRP control group scaled scores in comparison with the scaled scores achieved by experimental children in the other comparison programs, we see that the FDRP control children performed on many of the available subtest comparisons at a level superior to experimental children in comparison groups. There are seven subtests for which comparative data are available. In Figure 2 we note that on the Auditory Reception subtest and on the Grammatic Closure subtest, FDRP controls achieved a higher mean scaled score than experimentals in one of the two other programs with available comparison data. On the Visual Reception subtest, FDRP controls achieved a higher mean scaled score than experimentals in two of the three other programs with available comparison data. On the Auditory Reception subtest and on the Verbal Reception subtest, FDRP controls achieved a higher mean scaled score than experimentals in three of the other four programs with available comparison data. Only on one subtest, Visual Closure, did FDRP controls achieve a lower mean scaled score than that of the one other experimental comparison group available.

Table 3 illustrates the consequences of superior control group functioning on different scores calculated between control and experimental groups for each program.

 Insert Table 3 about here

We have subtracted the scaled scores achieved by control group children from the experimental scores of children in each project separately for each subtest. Data from only five of the comparison groups were used since experimental scores were only available in the other groups for experimental children well beyond the age of 4-0 to 5-0, the age range with which this study is concerned. It will be noted that in the Milwaukee project differences between experimental and control children were uniformly greater for all subtests than in any other project. By contrast, differences between experimental and control children in the FDRP project on each subtest were very modest, although uniformly in favor of the experimental group. Differences in the Durham project scores are difficult to interpret, since these scores do not represent experimental-control differences but represent pre- and post-intervention scores of the same children. The mixed direction of the calculated pre-post differences might lead one to infer that intervention had both detrimental and positive effects on the ITPA subscale performances of the children involved in the Durham project.

Discussion and Conclusions

The above results indicated that control children from a low-education, low-income population, where vulnerability in language areas is considered particularly likely, performed at satisfactory levels on subtests of a

nationally standardized language assessment instrument, the Illinois Test of Psycholinguistic Ability. These outcomes were associated with a carefully and consistently applied method of optimizing test conditions for children in both the experimental and control groups of an intervention program. We are thus led to reconsider some of the assumptions and expectations of intervention projects and their testing programs.

What are the implications of being below the national norms for the majority of control group children in intervention projects--children who are in many cases being tested in ways which we suspect are less than optimal? If we continue a) to use control groups such as those reported on here, and b) we have the children tested by conventionally trained testers, often ~~graduate students~~ with little training in sensitivity to children from different cultural and economic groups, and c) we also expect to get control group scores below the national norms, and d) indeed we get such scores, then we are creating stereotypic sets for these groups. These stereotypic sets might be inaccurate, and there is the possibility that we ourselves by our own testing methods are helping to create those stereotypic sets.

Traditionally, intervention studies have always been judged by comparisons of experimental and control groups. Let us take the hypothetical case of an intervention program where the post-test mean score for the experimental group on an outcome measure is 90 and for the control group 70. Is that program of value if the control group mean score is "really" 90 also. That is, under optimal testing conditions both groups would have scored approximately the same. Perhaps we have to re-examine the upper yield of some of our programs and intervention models. The maximal achievements of some models may in reality not be where we would want them to be.

Another point is that we may need to look at new ways to evaluate the impact of intervention programs. Is it possible that the old control-versus-experimental group model is not the most appropriate way to look at the difference a program has made in the achievements and performance of the children who have been involved. What about regional differences? Should we be impressed by significant E vs. C group differences if control children in a given region tend to perform quite poorly on a test? Our experience in this study suggests that we should keep our primary focus on program goals rather than on significant group differences where control group scores are regionally very low. In such cases perhaps we should look again at the experimental group scores and relate them to expected outcomes based on national norms rather than simply compare experimental scores with control group scores which reflect "significantly poorer" performance. In this way a more realistic assessment of the effects of a given program can be made.

Finally, we need to make recommendations for new styles of testing. We need a new level of quality. In addition, we need better techniques of quality control. Both of these are necessary to ensure uniformly sustained better quality of testing in our programs. One of the difficulties may lie in the fact that in addition to the procedures and directions given in testing manuals, there is a great deal of variance in the behaviors of testers--variance which is not addressed in the manuals. Personal styles can affect testing outcomes. Another point is that the manual's definition of a "trial" has always focused on what the tester is to do, almost in exclusion of what the child is doing. The definition of a "trial" should be that the child actually tries and not that the examiner simply makes a presentation that might not be adequately received by the child.

References

- Anastasiow, N. J., Stedman, D. J., & Spaulding, R. J. Language and reading achievement among disadvantaged nursery and primary grade children. Unpublished manuscript, Duke University, 1968.
- Baratz, S., & Baratz, J. Early childhood intervention: The social science base of institutional racism. Harvard Educational Review, 1970, 40 (1).
- Bereiter, C., & Engelmann, S. Teaching disadvantaged children in the preschool. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.
- Butterfield, E. C., & Zigler, E. Preinstitutional social deprivation and IQ changes in institutionalized retarded children. Journal of Abnormal Psychology, 1970, 75, 83-89.
- Cattell, P. Infant intelligence scale. New York: Psychological Corporation, 1960.
- Connors, K., & Eisenberg, L. The effect of teacher behavior on intelligence in Operation Head Start children. Unpublished manuscript, Johns Hopkins School of Medicine, Baltimore, Maryland, 1966.
- Costello, J., & Martin, J. Evaluation report: The first two years, Waves I & II, 1966-1967. Evaluation report: The last two years, Waves III & IV, 1967-69. Unpublished manuscript, Preschool Project, Institute for Juvenile Research, Chicago, Illinois.
- Garber, H. Personal communication, February, 1975.
- Heber, R., Garber, H., Harrington, S., Hoffman, C., & Falander, C. Rehabilitation of families at risk for mental retardation. Progress Report. Madison, Wisconsin: Rehabilitation Research and Training Center in Mental Retardation, University of Wisconsin, 1972.

- Karnes, M. B., Hodgins, A. S., Teska, J. A., & Kirk, S. A. Research and development program on preschool disadvantaged children; Investigation of classroom and at-home interventions. Final report. Washington, D. C.: United States Department of Health, Education, and Welfare, Office of Education, Bureau of Research, 1969.
- Kirk, S. A., McCarthy, J. J., & Kirk, W. D. Illinois Test of Psycholinguistic Abilities, Revised Edition. Urbana, Ill.: University of Illinois Press, 1968.
- Lally, J. R. Personal communication, Upstart Conference, Little Rock, Arkansas, 1970.
- Lally, J. R. Personal communication, Upstart Executive Committee meeting, Gainesville, Florida, 1972.
- Lally, J. R. The family development research program--a program for pre-natal, infant, and early childhood enrichment. Progress report. Office of Child Development, United States Department of Health, Education, and Welfare, 1974.
- Nedler, S., & Sebera, P. Intervention strategies for Spanish-speaking preschool children. Child Development, 1971, 42, 259-267.
- Palmer, F. H. Socioeconomic status and intellectual performance among Negro preschool boys. Developmental Psychology, 1970, 3 (1), 1-9.
- Quay, L. C. Language dialect, reinforcement, and the intelligence-test performance of Negro children. Child Development, 1971, 42, 5-15.
- Resnick, M. B., Weld, G. L., & Lally, J. R. Verbalizations of environmentally deprived two-year-olds as a function of the presence of a tester in a standardized test situation. Paper presented at a meeting of the American Educational Research Association, Chicago, 1969.

Sacks, E. L. Intelligence scores as a function of experimentally established social relationships between child and examiner. Journal of Abnormal and Social Psychology, 1952, 47, 354-458.

Valentine

Van de Riet, V. A sequential approach to early childhood education, Phase III. Grant report: submitted to the Office of Child Development, Department of Health, Education, & Welfare, January, 1972.

Yarrow, L. J. Interviewing children. In P. H. Mussen (Ed.), Handbook of research methods in child development. New York: John Wiley, 1960.

Zigler, E., Abelson, W. D., & Seitz, V. Motivational factors in the performance of economically disadvantaged children on the Peabody Picture Vocabulary Test. Child Development, 1973, 44 (2), 294-303.

Zigler, E., & Butterfield, E. C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, 1968, 39, 1-14.

Zigler, E., Butterfield, E. C., & Capobianco, F. Institutionalization and the effectiveness of social reinforcement: a five- and eight-year follow-up study. Developmental Psychology, 1970, 3 (2), 255-262.

Zigler, E., Hodgden, L., & Stevenson, H. W. The effect of support and non-support on the performance of normal and feeble-minded children: Journal of Personality, 1958, 26, 106-122.

Description of ITPA Subtests^a for Which Scores are Compared

Auditory Reception (Auditory Decoding)	This test assesses the ability of a child to understand verbally-presented increasingly difficult material. The response required is a simple head shake or nod or Yes-No answer. Sample question, "Do dials yawn?"
Visual Reception (Visual Decoding)	This test measures a child's ability to understand the meaning of a visual symbol. A stimulus picture X is on one page, and 4 response pictures are on another page. The child is to find the response picture conceptually similar to X. For example, X might be one kind of table.
Auditory-Vocal Association	The organizing process of using language symbols meaningfully is assessed in this sentence completion test which presents increasingly difficult analogy items, such as "Daddy is big; baby is _____."
Visual-Motor Association	This is a picture association test. A single stimulus picture is surrounded by 4 response pictures, one of which is associated with the stimulus pictures. The child has to choose a response picture most closely related to the stimulus picture; such as a sock, related to a stimulus picture of a shoe.
Verbal Expression (Vocal Encoding)	This test assesses the ability of a child to express as many of his own ideas and descriptions about a familiar object, such as a ball, as he can. He is asked, "Tell me all about this."
Manual Expression (Motor Encoding)	This test asks for gestural manipulations by a child to assess his understanding. He is required to pantomime actions appropriate to each pictured object, such as a telephone.
Grammatical Closure	The child's ability to use grammatic forms of Standard English is assessed. Syntax and grammatic inflections such as plurals and past tense forms are assessed in conjunction with pictured items. Example, "Here is a knife; here are two _____."
Visual Closure	This test assesses a child's ability in 30 seconds to detect or identify common objects which are each partially hidden in a complex pictorial scene.
Auditory Sequential Memory	This test is an assessment of the child's ability to reproduce from memory sequences of digits increasing in length from two to eight digits.
Visual Sequential Memory	This test assesses whether the child can, with chips of figures, create increasingly lengthy visual patterns identical in sequence to nonmeaningful figures presented for 5 seconds prior to the test.

^a The subtest name in parentheses refers to the name used in the 1961 version of the ITPA. The other name refers to the same subtest in the 1968 version of the ITPA.

Table 2

ITPA Scaled Scores Achieved by Control and Experimental Children
in Seven Intervention Projects

Group	Mean Age	N	Aud. Recep	Vis. Recep	Aud. Assoc	Vis. Assoc	Verb. Exp	Man. Exp	Gram. Clos	Vis. Clos	Aud. Seq	Vis. Seq
Family Development Research Program												
Exp.	4-0	58	36.4	41.0	37.5		43.3	50.2	34.8	40.1		
Controls	4-0	34	34.1	39.3	33.7		40.4	47.3	34.2	38.2		
Durham EIP												
Exp.	5-0	16	31.3	38.5	29.7	35.3	33.6	40.3	24.8		41.2	30.5
Controls												
Group 1	4-6	7										
Group 3	4-4.9	9										
Total	4-5.4	16	33.0	36.0	33.5	34.0	30.5	37.5	28.0		39.0	34.0
Karnes'												
Programs												
Exp.	Too											
Controls	Old											
Trad.	4-5.4	25										
Com.	4-1.3	16										
Montes	4-2.2	13										
Amelior.	4-4.6	24										
Direct	4-3.1	23										
Total	4-3.6	101	35.0	34.5	30.0		29.5	34.0	30.0			
Milwaukee												
Project												
(Heber)												
Exp.	4-7	18	36.8	46.1	40.7	41.1	42.3	43.9	42.4	43.5	40.6	39.9
Controls	4-7	18	28.8	37.8	29.4	31.8	34.4	38.2	28.7	31.8	33.1	29.3
Learning to												
Learn												
(Sprigle)												
Exp.	5-0	22		33.3	29.5		38.0					36.6
Controls	4-10	20		29.0	26.0		30.5					31.0

Table 2 (cont.)

Group	Mean Age	N	Aud. Recep	Vis. Recep.	Aud. Assoc.	Vis. Assoc.	Verb. Exp.	Man. Exp.	Gram. Clos.	Vis. Clos.	Aud. Seq.	Vis. Seq.
Bereiter- Engelmann Program Exp.	Too Old											
Controls	4-4	15			24.0		30.0		25.0			
Horner Preschool Program (Costello) Exp.	4-11.7	150 ^a 94 ^b			25.2		30.7					
Controls	4-5.2	132 ^c 67 ^d			26.0		26.5					

^a N for Auditory Association^b N for Verbal Expression^c N for Auditory Association^d N for Verbal Expression

Table 3

Differences Between Experimental and Control ITPA Scaled Scores
of Children Within Each Program Group

Group	Aud Recp.	Vis Recp	Aud Assoc	Vis Assoc	Verb Exp	Man Exp	Gram Clos	Vis Clos	Aud Seq	Vis Seq
Family Development Research Program	2.3	1.7	3.8		2.9	2.9	.6	1.9		
Durham EIP ^b	-1.7	2.5	-3.8	1.3	3.3	2.8	-3.2		2.2	-3.5
Milwaukee Project (Heber)	8.0	8.3	11.3	9.3	7.9	5.7	13.7	11.7	7.5	10.6
Learning To Learn (Sprigle)		4.3	3.5		7.5					5.6
Horner Preschool Program (Costello)			-1.8		4.2					

^a Ages of children in all groups range between 4-0 and 5-0.

^b In this analysis experimental children's pretest scores serve as control scores.

REPRESENTATIONAL

AUTOMATIC

RECEPTION

ASSOCIATION

EXPRESSION

CLOSURE

SEQ. MEMORY

AUD

VIS

AUD

VIS

VERB

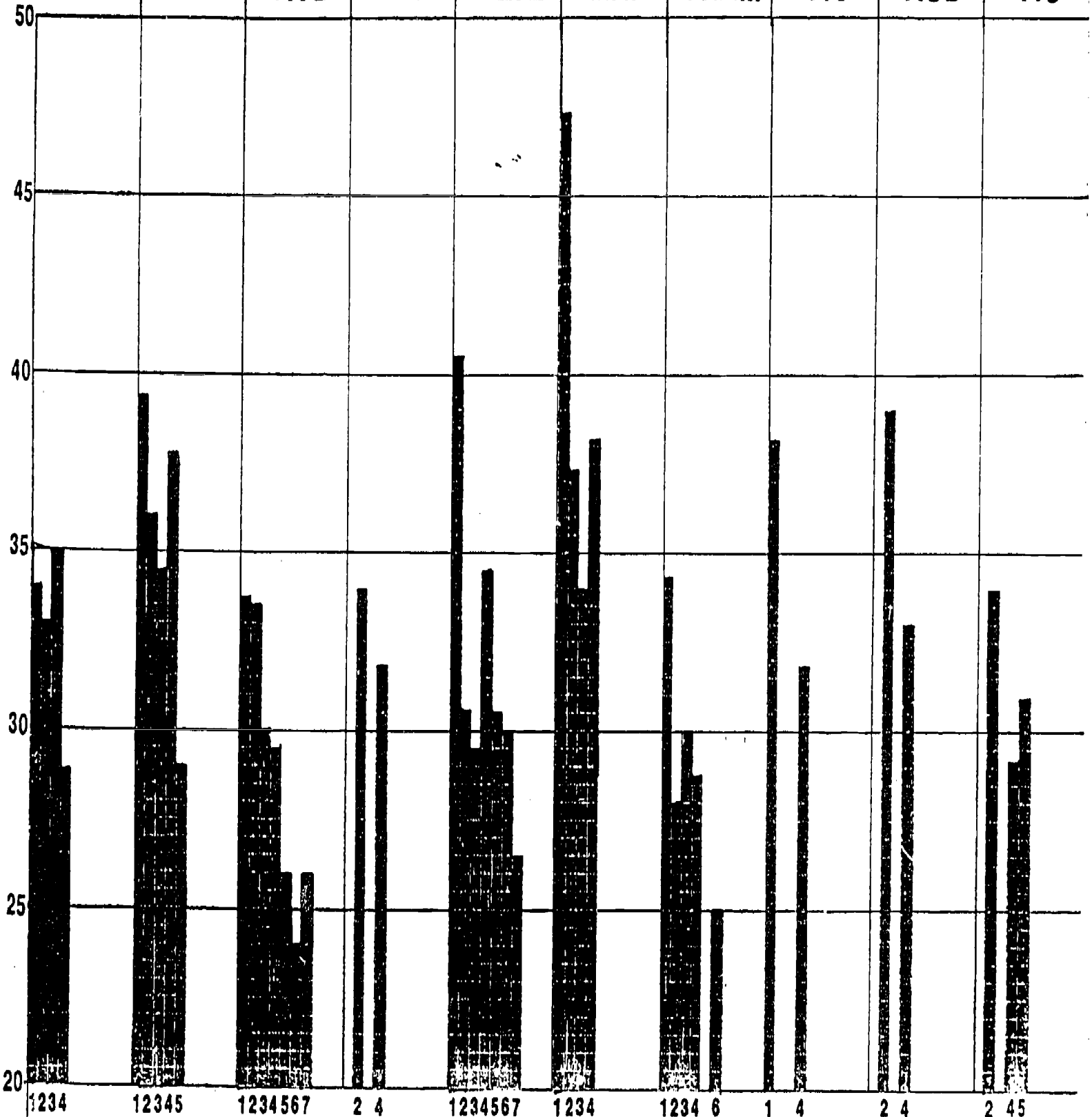
MAN

GRAM

VIS

AUD

VIS



1 FAMILY DEVELOPMENT RESEARCH PROGRAM

3 KARNES' PROGRAMS

5 LEARNING TO LEARN

ERIC HAM EDUCATION IMPROVEMENT PROJECT

4 MILWAUKEE PROJECT

6 BEREITER-ENGELMANN PROGRAM

7 HORNER PRESCHOOL PROGRAM

REPRESENTATIONAL

AUTOMATIC

RECEPTION

ASSOCIATION

EXPRESSION

CLOSURE

SEQ. MEMORY

AUD

VIS

AUD

VIS

VERB

MAN

GRAM

VIS

AUD

VIS

50

45

40

35

30

25

20

1234

28 12345

123456

a 34

123456

1234

1234

12 4

a 34

a 345

1 F.D.R.P. CONTROLS

2 MILWAUKEE PROJECT
3 EARLY DEVELOPMENT RESEARCH PROGRAM
4 LEARNING TO LEARN
5 HONOR PRESCHOOL PROGRAM

4 MILWAUKEE PROJECT

5 LEARNING TO LEARN

6 HORNER PRESCHOOL PROGRAM

^a Subtest not administered to either FDRP control or experimental children.

Appendix

Sample Profiles of Control Group Mothers in the
Family Development Research Program

4274 is a pretty 25 year old mother of three children who is separated from her husband. She works at a downtown store as a waitress and receives some support money from the childrens' father. She has attended ABC but has not received her high school diploma. She has high hopes for her children and would like to move from Syracuse. She would like to become a cosmologist some day. Occasionally, she works as an expensive party prostitute. Her relationship with her exhusband is sometimes violent and the children are very aware of this.

4205 is 19 years old and the mother of three children. She is single and is an inexpensive prostitute. She has abandoned her children and she comes home only when she is pregnant. The grandmother is now the chief caregiver since the state took the three children away from 4205. The extended family setting has seven children in all. 4205 has a drug problem in addition to frequent births and police troubles. She was enrolled at the Adult Basic Learning Center (ABC) trying to get a high school equivalency diploma. Family income continues to be mostly through social services and welfare benefits.

4308 is married with three children. At 20 she is still unable to cope with everyday living. She is very lethargic, unkempt, and looks and says she feels unhealthy. Her husband is unable to read and write well, The maternal grandmother over-sees the whole family unit. 4308 has had three miscarriages in the past 18 months and appears to be in very poor health. The grandmother is the chief caregiver with the children. 4308 lacks motivation even to dress herself in the morning. She has attended ABC program but as yet has no high school diploma. Her sole source of income is welfare. Her husband is often absent from the home and rarely works since he lacks skills.